

STIC-Biotech/ChemLib

170937

From: Russel, Jeffrey
Sent: Tuesday, November 08, 2005 9:42 AM
To: STIC-Biotech/ChemLib
Subject: Database Search Request

RECEIVED
NOV -8 2005
STIC/BIOTECH DIVISION
(STIC)

Requester:
Jeffrey Russel (TC1600)
Art Unit:
1654
Employee Number:
62785
Office Location:
REM 3D19
Phone Number:
571-272-0969
Mailbox Number:
REM 3C18

Checked
JFR
11-22-2005

Case serial number:
10/789,494
Class / Subclass(es):
NA
Earliest Priority Filing Date:
NA
Format preferred for results:
Diskette

Search Topic Information:

Please search SEQ ID NO:14 is the US patent application sequence databases (pending, published, and issued) and in Geneseq/Uniprot/PIR.
Thank you.

Special Instructions and Other Comments:

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2005, 21:47:02 ; Search time 163 Seconds
(without alignments)
68.810 Million cell updates/sec

Title: US-10-789-494B-14
Perfect score: 165
Sequence: 1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	146	88.5	219	5	AAM50040	Aam50040 N. clavip
2	146	88.5	264	5	AAM50048	Aam50048 N. clavip
3	64	38.8	345	8	ADS96556	Ads96556 Drosophil
4	64	38.8	378	4	ABB66461	Abb66461 Drosophil
5	60	36.4	89	7	ADD26163	Add26163 Protein r
6	57.5	34.8	248	4	ABB59150	Abb59150 Drosophil
7	57	34.5	134	6	ABU09590	Abu09590 Tick infe
8	57	34.5	154	2	AAy13500	Aay13500 Tissue ce
9	57	34.5	154	5	AAM50381	Aam50381 Tick ceme

10	57	34.5	154	6	ABU09592	Abu09592	Tick infe
11	55	33.3	110	6	ABU97112	Abu97112	Recombina
12	55	33.3	253	6	ABU97154	Abu97154	Recombina
13	55	33.3	301	4	ABB69239	Abb69239	Drosophil
14	55	33.3	352	2	AAW22358	Aaw22358	S. pneumo
15	55	33.3	378	6	ABU00882	Abu00882	S. pneumo
16	55	33.3	378	6	ABP81556	Abp81556	Streptoco
17	55	33.3	378	8	ADK48132	Adk48132	Streptoco
18	55	33.3	379	8	ADR94852	Adr94852	Novel S.
19	55	33.3	470	2	AAW72016	Aaw72016	HSV-2 str
20	54.5	33.0	508	8	ABM84985	Abm84985	Human dia
21	54.5	33.0	508	8	ABM84986	Abm84986	Human dia
22	54.5	33.0	522	2	AAW36052	Aaw36052	Human occ
23	54.5	33.0	522	2	AAW34638	Aaw34638	Human occ
24	54.5	33.0	522	3	AAB35731	Aab35731	Human occ
25	54.5	33.0	522	6	ABJ37076	Abj37076	Human bre
26	54.5	33.0	522	7	ADD46545	Add46545	Human Pro
27	54.5	33.0	522	8	ADI47189	Adi47189	Human occ
28	54	32.7	314	8	ADR21277	Adr21277	Streptomy
29	54	32.7	330	3	AAG06302	Aag06302	Arabidops
30	54	32.7	336	3	AAG06301	Aag06301	Arabidops
31	54	32.7	344	3	AAG06300	Aag06300	Arabidops
32	54	32.7	344	8	ADN72757	Adn72757	Thale cre
33	54	32.7	416	4	ABB67901	Abb67901	Drosophil
34	54	32.7	419	7	ABO74499	Abo74499	Pseudomon
35	53.5	32.4	180	6	ABG73439	Abg73439	Common du
36	53	32.1	281	5	AAG77977	Aag77977	Human NK-
37	53	32.1	301	5	AAG77976	Aag77976	Human NK-
38	53	32.1	367	4	ABG26521	Abg26521	Novel hum
39	53	32.1	701	8	ADJ34790	Adj34790	Xylanase
40	52.5	31.8	122	4	ABB69531	Abb69531	Drosophil
41	52.5	31.8	194	4	AAG89977	Aag89977	C glutami
42	52	31.5	298	6	ABG73438	Abg73438	Common du
43	52	31.5	458	2	AAR98744	Aar98744	Nuclear e
44	52	31.5	672	8	ADS26510	Ads26510	Bacterial
45	52	31.5	676	8	ADS27253	Ads27253	Bacterial

ALIGNMENTS

RESULT 1

AAM50040

ID AAM50040 standard; protein; 219 AA.

XX

AC AAM50040;

XX

DT 18-SEP-2002 (first entry)

XX

DE N. clavipes spidroin synthetic homologue FA2 protein.

XX

KW Spidroin; spider; silk; fibre; film; membrane; wound; filter; FA2.

XX

OS Synthetic.

XX

PN DE10113781-A1.

XX

PD 13-DEC-2001.
 XX
 PF 21-MAR-2001; 2001DE-01013781.
 XX
 PR 09-JUN-2000; 2000DE-01028212.
 PR 24-OCT-2000; 2000DE-01053478.
 XX
 PA (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
 XX
 PI Scheller J, Conrad U, Grosse F, Guehrs K;
 XX
 DR WPI; 2002-123561/17.
 DR N-PSDB; ABL61041.
 XX
 PT New DNA encoding synthetic spider silk protein, useful e.g. for closing
 PT wounds, comprises modules that encode repeating units of spidroin
 PT proteins.
 XX
 PS Claim 22; Page 46-47; 88pp; German.
 XX
 CC This invention describes a novel DNA sequence, encoding a synthetic
 CC spider silk protein, comprising modules, each comprising a group of
 CC sequentially arranged oligonucleotides, each oligonucleotide encoding a
 CC repeating unit of a spidroin protein. The synthetic protein has at least
 CC 84% homology with the Nephila clavipes spidroin protein and is used to
 CC produce synthetic fibres, films and/or membranes, particularly: (i) for
 CC medical use, especially to close wounds and/or to support or cover
 CC artificial organs; (ii) as adhesion surfaces for culturing cells; and
 CC (iii) as filters. The synthetic proteins are very similar to native
 CC spider silk proteins; can be prepared on a large scale and can be spun to
 CC fibres with excellent mechanical properties (strength and elasticity).
 CC Also they retain water solubility after long-term boiling in aqueous
 CC solutions and since they are also soluble in organic solvents but
 CC precipitated at high salt concentration, they are easily extracted and
 CC purified. The modular construction of the invention facilitates
 CC incorporation of additional peptide-encoding sequences, e.g. to simplify
 CC purification or modulate solubility. This sequence represents the
 CC synthetic N. clavipes spidroin-1 homologue FA2 described in the invention
 XX
 SQ Sequence 219 AA;

Query Match 88.5%; Score 146; DB 5; Length 219;
 Best Local Similarity 89.7%; Pred. No. 2.6e-11;
 Matches 26; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29
 |||||:|||||:|||||
 Db 186 GSSGFGPYVANGGYSGYEYAWSSKSDFET 214

RESULT 2
 AAM50048
 ID AAM50048 standard; protein; 264 AA.
 XX
 AC AAM50048;
 XX
 DT 18-SEP-2002 (first entry)

XX
DE N. clavipes spidroin synthetic homologue FA2 protein #2.
XX
KW Spidroin; spider; silk; fibre; film; membrane; wound; filter; FA2.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1. .28
FT /label= LeB4_signal_peptide
FT Protein 29. .247
FT /note= "Synthetic spidroin and fibre protein homologue
FT FA2"
FT Region 248. .260
FT /note= "c-myc-tag"
FT Domain 261. .264
FT /note= "ER retention signal"
XX
PN DE10113781-A1.
XX
PD 13-DEC-2001.
XX
PF 21-MAR-2001; 2001DE-01013781.
XX
PR 09-JUN-2000; 2000DE-01028212.
PR 24-OCT-2000; 2000DE-01053478.
XX
PA (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX
PI Scheller J, Conrad U, Grosse F, Guehrs K;
XX
DR WPI; 2002-123561/17.
XX
PT New DNA encoding synthetic spider silk protein, useful e.g. for closing
PT wounds, comprises modules that encode repeating units of spirodoin
PT proteins.
XX
PS Example 1; Fig 10B; 88pp; German.
XX
CC This invention describes a novel DNA sequence, encoding a synthetic
CC spider silk protein, comprising modules, each comprising a group of
CC sequentially arranged oligonucleotides, each oligonucleotide encoding a
CC repeating unit of a spidroin protein. The synthetic protein has at least
CC 84% homology with the Nephila clavipes spidroin protein and is used to
CC produce synthetic fibres, films and/or membranes, particularly: (i) for
CC medical use, especially to close wounds and/or to support or cover
CC artificial organs; (ii) as adhesion surfaces for culturing cells; and
CC (iii) as filters. The synthetic proteins are very similar to native
CC spider silk proteins; can be prepared on a large scale and can be spun to
CC fibres with excellent mechanical properties (strength and elasticity).
CC Also they retain water solubility after long-term boiling in aqueous
CC solutions and since they are also soluble in organic solvents but
CC precipitated at high salt concentration, they are easily extracted and
CC purified. The modular construction of the invention facilitates
CC incorporation of additional peptide-encoding sequences, e.g. to simplify
CC purification or modulate solubility. This sequence represents a construct
CC composed of the LeB4 signal peptide, N. clavipes spidroin-1 and fibre

CC protein synthetic homologue FA2, a c-Myc-tag and an endoplasmic reticulum
CC (ER)-retention signal described in the invention
XX
SQ Sequence 264 AA;

Query Match 88.5%; Score 146; DB 5; Length 264;
Best Local Similarity 89.7%; Pred. No. 3.1e-11;
Matches 26; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29
|||||||:|||||||:|||
Db 213 GSSGFGPYVANGGYSGYEYAWSSKSFET 241

Search completed: November 8, 2005, 22:03:28
Job time : 165 secs

OM protein - protein search, using sw model

Run on: November 8, 2005, 21:47:56 ; Search time 39 Seconds
(without alignments)
71.546 Million cell updates/sec

Title: US-10-789-494B-14
Perfect score: 165
Sequence: 1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Match	Length			
1	64	38.8	308	2	B47369	RNA-binding protei
2	64	38.8	321	2	A47369	RNA-binding protei
3	64	38.8	345	1	B41732	heterogeneous nucl
4	60	36.4	89	2	T25923	hypothetical prote
5	60	36.4	353	1	S56750	single stranded D
6	58	35.2	152	2	T07858	glycine-rich prote
7	55	33.3	378	2	D95060	dnaJ protein [impo
8	55	33.3	636	2	F69027	cleavage and polya
9	54.5	33.0	522	2	G02533	occludin - human
10	54	32.7	142	2	C33910	sal homeotic prote
11	54	32.7	178	2	T19215	hypothetical prote
12	54	32.7	336	2	T05538	hypothetical prote
13	54	32.7	401	2	C83109	probable transport

14	54	32.7	509	2	T40835	hypothetical prote
15	53.5	32.4	406	2	G71404	probable ribonucle
16	53	32.1	422	2	T51199	hypothetical prote
17	53	32.1	1287	2	I46032	nuclear DNA helica
18	53	32.1	2639	2	T31328	fibroin - Chinese
19	52.5	31.8	139	2	T34244	hypothetical prote
20	52.5	31.8	287	2	D90540	glucokinase (gluco
21	52	31.5	59	2	H24802	cuticle protein 64
22	52	31.5	631	2	T13115	protein gp29 - pha
23	52	31.5	975	2	T16073	hypothetical prote
24	51.5	31.2	64	2	T21841	hypothetical prote
25	51.5	31.2	108	2	T26825	hypothetical prote
26	51.5	31.2	123	2	A69884	cell wall protein
27	51.5	31.2	363	2	S66727	hypothetical prote
28	51	30.9	284	2	T23158	hypothetical prote
29	51	30.9	300	2	JQ2220	hydroxyproline-ric
30	51	30.9	534	2	S62572	hypothetical prote
31	50.5	30.6	159	2	C49773	ecdysone-dependent
32	50.5	30.6	161	2	B42627	cement precursor p
33	50.5	30.6	345	2	B97066	aldose-1-epimerase
34	50.5	30.6	605	2	JH0638	alpha-amylase (EC
35	50.5	30.6	1324	2	T17468	peptide-synthetase
36	50	30.3	72	2	E89016	protein B0213.2 [i
37	50	30.3	88	2	A75340	hypothetical prote
38	50	30.3	128	2	JQ1002	keratin, claw - ch
39	50	30.3	139	2	T33968	hypothetical prote
40	50	30.3	212	2	T10553	hypothetical prote
41	50	30.3	227	2	T15772	hypothetical prote
42	50	30.3	629	2	T06675	hypothetical prote
43	49.5	30.0	64	2	T27944	hypothetical prote
44	49.5	30.0	225	2	A86903	hypothetical prote
45	49.5	30.0	371	2	I46089	thyroid transcript

ALIGNMENTS

RESULT 1

B47369

RNA-binding protein (alternatively spliced) SqdB - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 16-Aug-2004

C;Accession: B47369

R;Kelley, R.L.

Genes Dev. 7, 948-960, 1993

A;Title: Initial organization of the *Drosophila* dorsoventral axis depends on an RNA-binding protein encoded by the squid gene.

A;Reference number: A47369; MUID:93279471; PMID:7684991

A;Accession: B47369

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-308 <KEL>

A;Cross-references: UNIPROT:Q08473; GB:S62100; NID:g385453; PIDN:AAB26989.1; PID:g385455

A;Note: sequence extracted from NCBI backbone (NCBIN:132997, NCBIN:132999, NCBIP:133001)

C;Genetics:

A;Gene: FlyBase:sqd

A;Cross-references: FlyBase:FBgn0003498

C;Superfamily: ribonucleoprotein repeat homology

F;57-123/Domain: ribonucleoprotein repeat homology <RRM1>

F;137-203/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 38.8%; Score 64; DB 2; Length 308;
Best Local Similarity 47.6%; Pred. No. 0.48;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 GSSGFGPYVAHGGYSGYEYAW 21
|:|:| | | | |:|:|:| :
Db 265 GAGGYGDYYAGGYNGYDYG 285

RESULT 2

A47369

RNA-binding protein (alternatively spliced) SqdA - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C;Accession: A47369; C41732

R;Kelley, R.L.

Genes Dev. 7, 948-960, 1993

A;Title: Initial organization of the *Drosophila* dorsoventral axis depends on an RNA-binding protein encoded by the squid gene.

A;Reference number: A47369; MUID:93279471; PMID:7684991

A;Accession: A47369

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-321 <KEL>

A;Cross-references: UNIPROT:Q08473; UNIPROT:Q8MSY1; GB:S61875; NID:g385452; PIDN:AAB26988.1; PID:g385454

A;Note: sequence extracted from NCBI backbone (NCBIN:132997, NCBIP:133000)

R;Matunis, E.L.; Matunis, M.J.; Dreyfuss, G.

J. Cell Biol. 116, 257-269, 1992

A;Title: Characterization of the major hnRNP proteins from *Drosophila melanogaster*.

A;Reference number: A41732; MUID:92112968; PMID:1730754

A;Accession: C41732

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-168,'F',170-321 <MAT>

A;Cross-references: GB:X62637; GB:S76630; NID:g11037; PIDN:CAA44503.1; PID:g11038

A;Note: sequence extracted from NCBI backbone (NCBIN:76630, NCBIP:76631)

C;Genetics:

A;Gene: FlyBase:sqd

A;Cross-references: FlyBase:FBgn0003498

C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat homology

F;57-123/Domain: ribonucleoprotein repeat homology <RRM1>

F;137-203/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 38.8%; Score 64; DB 2; Length 321;
Best Local Similarity 47.6%; Pred. No. 0.5;

Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```
Qy      1 GSSGFGPYVAHGGYSGYEYAW 21
          |: |:| | | | |:| |:| :
Db      265 GAGGYGDYYAGGYNGYDYG 285
```

Search completed: November 8, 2005, 22:04:12
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2005, 22:00:52 ; Search time 169 Seconds
(without alignments)
87.872 Million cell updates/sec

Title: US-10-789-494B-14
Perfect score: 165
Sequence: 1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	165	100.0	5263	1	FBOH_BOMMO	P05790 bombyx mori
2	64	38.8	344	1	SQD_DROME	Q08473 drosophila
3	61.5	37.3	492	2	Q7ZUE7	Q7zue7 brachydanio
4	60	36.4	89	2	Q23052	Q23052 caenorhabdi
5	60	36.4	166	2	Q8MV46	Q8mv46 trypanosoma
6	60	36.4	166	2	Q7Z1G9	Q7z1g9 trypanosoma
7	59	35.8	693	2	Q6K5F8	Q6k5f8 oryza sativ
8	58.5	35.5	157	2	Q6F4A0	Q6f4a0 streptomyce
9	58	35.2	152	2	Q41349	Q41349 lycopersico
10	58	35.2	393	2	Q6HQU5	Q6hqu5 bacillus an
11	58	35.2	393	2	Q9L4R8	Q9l4r8 bacillus ce
12	58	35.2	393	2	Q9XBH5	Q9xbh5 bacillus ce
13	58	35.2	393	2	Q72XQ8	Q72xq8 bacillus ce
14	58	35.2	393	2	Q815F2	Q815f2 bacillus ce
15	58	35.2	393	2	Q81X10	Q81x10 bacillus an

16	58	35.2	393	2	Q6HB83	Q6hb83	bacillus th
17	57.5	34.8	212	2	Q8IRH6	Q8irh6	drosophila
18	57.5	34.8	242	2	Q8MZ31	Q8mz31	drosophila
19	57.5	34.8	242	2	Q9W0H1	Q9w0h1	drosophila
20	57	34.5	132	2	Q7Q1T7	Q7qlt7	anopheles g
21	57	34.5	154	2	Q8T6I1	Q8t6i1	rhipicephal
22	57	34.5	346	2	Q82GU6	Q82gu6	streptomyce
23	57	34.5	393	2	Q631E8	Q631e8	bacillus ce
24	57	34.5	640	2	Q84XZ4	Q84xz4	triticum ae
25	56.5	34.2	242	2	Q6P642	Q6p642	xenopus tro
26	56.5	34.2	287	2	Q17200	Q17200	bombyx mori
27	56.5	34.2	303	2	Q17201	Q17201	bombyx mori
28	56	33.9	299	2	Q74D41	Q74d41	geobacter s
29	56	33.9	409	2	Q673W4	Q673w4	mus musculu
30	56	33.9	432	1	K3L1_MOUSE	P83555	mus musculu
31	56	33.9	432	2	Q673W3	Q673w3	mus musculu
32	55.5	33.6	362	2	Q6Z8U4	Q6z8u4	oryza sativ
33	55.5	33.6	381	2	Q9GP09	Q9gp09	ixodes rici
34	55.5	33.6	464	2	Q7XDI5	Q7xdi5	oryza sativ
35	55.5	33.6	464	2	Q9FWK8	Q9fwk8	oryza sativ
36	55.5	33.6	500	2	Q6NX99	Q6nx99	brachydanio
37	55.5	33.6	1172	2	Q9LWY9	Q9lwy9	oryza sativ
38	55	33.3	75	2	Q8T3D9	Q8t3d9	caenorhabdi
39	55	33.3	109	2	Q7BKH6	Q7bkh6	gamma-prote
40	55	33.3	226	2	Q6NWF9	Q6nwf9	brachydanio
41	55	33.3	272	2	Q9VEI2	Q9vei2	drosophila
42	55	33.3	378	1	DNAJ_STRPN	P95830	streptococc
43	55	33.3	388	2	Q673W2	Q673w2	mus musculu
44	55	33.3	505	2	Q9U913	Q9u913	procambarus
45	55	33.3	636	2	O27271	O27271	methanobact

ALIGNMENTS

RESULT 1

FBOH_BOMMO

ID FBOH_BOMMO STANDARD; PRT; 5263 AA.

AC P05790; Q17220; Q26379;

DT 01-NOV-1988 (Rel. 09, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Fibroin heavy chain precursor (Fib-H) (H-fibroin).

GN Name=FIBH;

OS Bombyx mori (Silk moth).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;

OC Bombycidae; Bombyx.

OX NCBI_TaxID=7091;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20330362; PubMed=10871375; DOI=10.1093/nar/28.12.2413;

RA Zhou C.-Z., Confalonieri F., Medina N., Zivanovic Y., Esnault C.,

RA Yang T., Jacquet M., Janin J., Duguet M., Perasso R., Li Z.-G.;

RT "Fine organization of Bombyx mori fibroin heavy chain gene.";

RL Nucleic Acids Res. 28:2413-2419(2000).

RN [2]

RP SEQUENCE OF 1-168 FROM N.A.
 RX MEDLINE=80045039; PubMed=498286; DOI=10.1016/0092-8674(79)90075-8;
 RA Tsujimoto Y., Suzuki Y.;
 RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
 RT flanking, mRNA coding, entire intervening and fibroin protein coding
 RT regions.";
 RL Cell 18:591-600(1979).
 RN [3]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=79211211; PubMed=455439; DOI=10.1016/0092-8674(79)90018-7;
 RA Tsujimoto Y., Suzuki Y.;
 RT "Structural analysis of the fibroin gene at the 5' end and its
 RT surrounding regions.";
 RL Cell 16:425-436(1979).
 RN [4]
 RP PARTIAL SEQUENCE FROM N.A.
 RC STRAIN=Kinshu X Showa;
 RX MEDLINE=89094868; PubMed=3210244;
 RA Mita K., Ichimura S., Zama M., James T.C.;
 RT "Specific codon usage pattern and its implications on the secondary
 RT structure of silk fibroin mRNA.";
 RL J. Mol. Biol. 203:917-925(1988).
 RN [5]
 RP PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=94365842; PubMed=7916056;
 RA Mita K., Ichimura S., James T.C.;
 RT "Highly repetitive structure and its organization of the silk fibroin
 RT gene.";
 RL J. Mol. Evol. 38:583-592(1994).
 RN [6]
 RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
 RC STRAIN=J-139;
 RX MEDLINE=99296390; PubMed=10366732; DOI=10.1016/S0167-4838(99)00088-6;
 RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
 RA Takagi T., Mizuno S.;
 RT "Determination of the site of disulfide linkage between heavy and
 RT light chains of silk fibroin produced by Bombyx mori.";
 RL Biochim. Biophys. Acta 1432:92-103(1999).
 CC -!- FUNCTION: Forms the silk filament; a strong, inextensible,
 CC insoluble and chemically inert fibre.
 CC -!- SUBUNIT: Formed of two chains: heavy and light, that are linked by
 CC a disulfide bond. Heavy-light chain assembly is essential for the
 CC efficient intracellular transport and secretion of fibroin.
 CC -!- TISSUE SPECIFICITY: Produced exclusively in the posterior (PSG)
 CC section of silk glands.
 CC -!- DOMAIN: Composed of antiparallel beta sheets. The strands of the
 CC beta sheets run parallel to the fiber axis. Long stretches of silk
 CC fibroin are composed of microcrystalline arrays of (-Gly-Ser-Gly-
 CC Ala-Gly-Ala-)n interrupted by regions containing bulkier residues.
 CC The fiber is composed of microcrystalline arrays alternating with
 CC amorphous regions.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF226688; AAF76983.1; -.
DR EMBL; V00094; CAA23432.1; -.
DR EMBL; V00097; CAA23433.1; -.
DR EMBL; S74439; AAB31861.1; -.
DR EMBL; X13869; CAA32076.1; -.
DR EMBL; M35378; AAA27839.1; -.
DR EMBL; AB017362; BAA33147.1; -.
DR PIR; S01844; S01844.
KW Repeat; Signal; Silk.
FT SIGNAL 1 21 Potential.
FT CHAIN 22 5263 Fibroin heavy chain.
FT DOMAIN 149 5206 Highly repetitive.
FT DISULFID 5244 5244 Interchain (with light chain).
FT DISULFID 5260 5263
FT CONFLICT 10 10 C -> V (in Ref. 2).
SQ SEQUENCE 5263 AA; 391586 MW; 8EE11D3A0A47440E CRC64;

Query Match 100.0%; Score 165; DB 1; Length 5263;
Best Local Similarity 100.0%; Pred. No. 5.3e-12;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29
|||
Db 1228 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 1256

RESULT 2

SQD_DROME

ID SQD_DROME STANDARD; PRT; 344 AA.
AC Q08473; Q26273; Q8IH71; Q8INH1; Q8MSY1; Q9VFT5; Q9VFT6;
DT 01-FEB-1995 (Rel. 31, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE RNA-binding protein squid (Heterogeneous nuclear ribonucleoprotein 40)
DE (HNRNP 40).
GN Name=sqd; Synonyms=hrp40; ORFNames=CG16901;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A AND C), FUNCTION, AND SUBCELLULAR
RP LOCATION.
RC TISSUE=Ovary;
RX MEDLINE=93279471; PubMed=7684991;
RA Kelley R.L.;
RT "Initial organization of the Drosophila dorsoventral axis depends on
RT an RNA-binding protein encoded by the squid gene."
RL Genes Dev. 7:948-960(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM B AND C).
RC STRAIN=Canton-S; TISSUE=Embryo;
RX MEDLINE=92112968; PubMed=1730754; DOI=10.1083/jcb.116.2.257;

RA Matunis E.L., Matunis M.J., Dreyfuss G.;
 RT "Characterization of the major hnRNP proteins from *Drosophila*
 RT *melanogaster*.";
 RL J. Cell Biol. 116:257-269(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [4]
 RP GENOME REANNOTATION, AND ALTERNATIVE SPLICING.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a

RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC STRAIN=Berkeley; TISSUE=Embryo, and Head;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource.";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 RN [6]
 RP SEQUENCE OF 59-102 FROM N.A. (ISOFORMS A/B/C).
 RX MEDLINE=93109300; PubMed=8417324;
 RA Kim Y.J., Baker B.S.;
 RT "Isolation of RRM-type RNA-binding protein genes and the analysis of
 RT their relatedness by using a numerical approach.";
 RL Mol. Cell. Biol. 13:174-183(1993).
 CC -!- FUNCTION: This protein is a component of ribonucleosomes. Could be
 CC needed to organize a concentration gradient of a dorsalizing
 CC morphogen (Dm) originating in the germinal vesicle. At least one
 CC of the isoforms is essential in somatic tissues.
 CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. It is possible that
 CC some isoforms are found only in one of these locations.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Additional isoforms seem to exist;
 CC Name=B; Synonyms=SqdS, HRP40.2;
 CC IsoId=Q08473-1; Sequence=Displayed;
 CC Name=A; Synonyms=SqdA, HRP40.1;
 CC IsoId=Q08473-2; Sequence=VSP_005876;
 CC Name=C; Synonyms=SqdB;
 CC IsoId=Q08473-3; Sequence=VSP_005877;
 CC Name=D;
 CC IsoId=Q08473-4; Sequence=VSP_011797;
 CC Note=No experimental confirmation available;
 CC -!- MISCELLANEOUS: Female mutants are sterile and lay eggs that
 CC display only dorsal structures.
 CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S61875; AAB26988.1; -.
 DR EMBL; S62100; AAB26989.1; -.
 DR EMBL; S61875; AAB26989.1; JOINED.
 DR EMBL; X62637; CAA44503.1; -.
 DR EMBL; X62638; CAA44504.1; -.
 DR EMBL; AE003701; AAF54963.2; -.
 DR EMBL; AE003701; AAF54964.2; -.
 DR EMBL; AE003701; AAN13570.1; -.
 DR EMBL; AE003701; AAS65146.1; -.

DR EMBL; AY118501; AAM49870.1; -.
 DR EMBL; BT001384; AAN71139.1; -.
 DR EMBL; BT003283; AAO25040.1; -.
 DR EMBL; S51693; AAB24624.1; -.
 DR PIR; A47369; A47369.
 DR PIR; B41732; B41732.
 DR PIR; B47369; B47369.
 DR HSSP; Q14103; 1HD1.
 DR IntAct; Q8MSY1; -.
 DR FlyBase; FBgn0003498; sqd.
 DR GO; GO:0000785; C:chromatin; IDA.
 DR GO; GO:0016607; C:nuclear speck; IDA.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0008298; P:mRNA localization, intracellular; NAS.
 DR GO; GO:0006406; P:mRNA-nucleus export; NAS.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; RRM_1; 2.
 DR PROSITE; PS50102; RRM; 2.
 KW Alternative splicing; Nuclear protein; Repeat; Ribonucleoprotein;
 KW RNA-binding.
 FT DOMAIN 56 138 RNA-binding (RRM) 1.
 FT DOMAIN 136 213 RNA-binding (RRM) 2.
 FT DOMAIN 221 337 Gly-rich.
 FT VARSPLIC 1 166 Missing (in isoform D).
 FT /FTId=VSP_011797.
 FT VARSPLIC 286 344 DGYGYGGGFEGNGYGGGGGGMGGGRGGPRGGGGPKGGGGF
 FT NGGKQRGGGGRRQQRHQPYPY -> GKYNKQQSSAQNNYYNNNT
 FT SSNYHQNKNNNSNNYQQF (in isoform A).
 FT /FTId=VSP_005876.
 FT VARSPLIC 286 322 Missing (in isoform C).
 FT /FTId=VSP_005877.
 FT CONFLICT 84 84 S -> N (in Ref. 6).
 FT CONFLICT 169 169 F -> L (in Ref. 1).
 FT CONFLICT 305 305 G -> GG (in Ref. 2; CAA44504).
 SQ SEQUENCE 344 AA; 36184 MW; 68E84791A924EED4 CRC64;

 Query Match 38.8%; Score 64; DB 1; Length 344;
 Best Local Similarity 47.6%; Pred. No. 4.4;
 Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

 QY 1 GSSGFGPYVAHGGYSGYEYAW 21
 | : | : | | | : | : | :
 Db 265 GAGGYGDYYAGGYNGYDYG 285

Search completed: November 8, 2005, 22:11:07
 Job time : 172 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2005, 21:57:52 ; Search time 69 Seconds
(without alignments)
175.853 Million cell updates/sec

Title: US-10-789-494B-14
Perfect score: 165
Sequence: 1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867879 seqs, 418409474 residues

Total number of hits satisfying chosen parameters: 1867879

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep:*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query Match	Length	DB	ID	Description	
1	165	100.0	29	18	US-10-789-494B-4	Sequence 4, Appli	
2	165	100.0	29	18	US-10-789-494B-11	Sequence 11, Appl	
3	165	100.0	29	18	US-10-789-494B-13	Sequence 13, Appl	
4	165	100.0	29	18	US-10-789-494B-14	Sequence 14, Appl	
5	165	100.0	29	18	US-10-789-494B-60	Sequence 60, Appl	
6	158	95.8	29	18	US-10-789-494B-12	Sequence 12, Appl	
7	158	95.8	29	18	US-10-789-494B-15	Sequence 15, Appl	
8	158	95.8	29	18	US-10-789-494B-16	Sequence 16, Appl	
9	158	95.8	29	18	US-10-789-494B-17	Sequence 17, Appl	
10	158	95.8	29	18	US-10-789-494B-19	Sequence 19, Appl	
11	143.5	87.0	28	18	US-10-789-494B-18	Sequence 18, Appl	
12	134.5	81.5	32	18	US-10-789-494B-20	Sequence 20, Appl	
13	126.5	76.7	30	18	US-10-789-494B-10	Sequence 10, Appl	
14	64	38.8	378	20	US-11-097-143-26175	Sequence 26175, A	
15	61.5	37.3	177	16	US-10-425-115-193655	Sequence 193655,	
16	59.5	36.1	215	15	US-10-425-114-68305	Sequence 68305, A	
17	59.5	36.1	450	16	US-10-425-115-193654	Sequence 193654,	
18	59.5	36.1	454	16	US-10-767-701-45105	Sequence 45105, A	
19	59.5	36.1	478	15	US-10-425-114-58912	Sequence 58912, A	
20	59.5	36.1	480	15	US-10-425-114-61022	Sequence 61022, A	
21	59	35.8	693	16	US-10-437-963-115279	Sequence 115279,	
22	58.5	35.5	126	16	US-10-767-701-56707	Sequence 56707, A	
23	58	35.2	295	16	US-10-425-115-193656	Sequence 193656,	
24	57.5	34.8	248	20	US-11-097-143-4242	Sequence 4242, Ap	
25	57	34.5	64	18	US-10-492-072-20	Sequence 20, Appl	
26	57	34.5	133	18	US-10-492-072-21	Sequence 21, Appl	
27	57	34.5	134	14	US-10-280-114-13	Sequence 13, Appl	
28	57	34.5	154	14	US-10-226-489-16	Sequence 16, Appl	
29	57	34.5	154	14	US-10-280-114-17	Sequence 17, Appl	
30	57	34.5	154	18	US-10-492-072-12	Sequence 12, Appl	
31	57	34.5	154	18	US-10-492-072-16	Sequence 16, Appl	
32	57	34.5	346	14	US-10-156-761-11334	Sequence 11334, A	
33	57	34.5	645	16	US-10-739-930-10518	Sequence 10518, A	
34	56.5	34.2	449	15	US-10-424-599-285485	Sequence 285485,	
35	56.5	34.2	529	15	US-10-425-114-49406	Sequence 49406, A	
36	55.5	33.6	195	16	US-10-437-963-157867	Sequence 157867,	
37	55.5	33.6	362	16	US-10-437-963-112439	Sequence 112439,	
38	55.5	33.6	1447	16	US-10-437-963-114974	Sequence 114974,	
39	55	33.3	110	16	US-10-479-670-152	Sequence 152, App	
40	55	33.3	253	16	US-10-479-670-194	Sequence 194, App	
41	55	33.3	301	20	US-11-097-143-34509	Sequence 34509, A	
42	55	33.3	378	16	US-10-474-776-634	Sequence 634, App	
43	55	33.3	378	17	US-10-472-928-900	Sequence 900, App	
44	55	33.3	379	18	US-10-617-320-3487	Sequence 3487, Ap	
45	55	33.3	636	10	US-09-988-626-237	Sequence 237, App	

ALIGNMENTS

RESULT 1
US-10-789-494B-4

```
; Sequence 4, Application US/10789494B
; Publication No. US20050143296A1
; GENERAL INFORMATION:
; APPLICANT: TSUBOUCHI, Kozo
; APPLICANT: YAMADA, Hiromi
; TITLE OF INVENTION: EXTRACTION AND UTILIZATION OF CELL
; TITLE OF INVENTION: GROWTH-PROMOTING PEPTIDES FROM SILK PROTEIN
; FILE REFERENCE: OPS 635
; CURRENT APPLICATION NUMBER: US/10/789,494B
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: JP 2003-55048
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 4
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Bombyx mori
US-10-789-494B-4
```

```
Query Match          100.0%; Score 165; DB 18; Length 29;
Best Local Similarity 100.0%; Pred. No. 5.4e-15;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29
          |||||||||||||||||||||
Db      1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29
```

RESULT 14

US-11-097-143-26175

```
; Sequence 26175, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
```

; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26175
; LENGTH: 378
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-26175

Query Match 38.8%; Score 64; DB 20; Length 378;
Best Local Similarity 47.6%; Pred. No. 3.3;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GSSGFGPYVAHGGYSGYEYAW 21
|: |:| | | |:| |:| :
Db 322 GAGGYGDYYAGGYNGYDYG 342

Search completed: November 8, 2005, 22:06:15
Job time : 70 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2005, 21:55:56 ; Search time 42 Seconds
(without alignments)
51.543 Million cell updates/sec

Title: US-10-789-494B-14
Perfect score: 165
Sequence: 1 GSSGFGPYVAHGGYSGYEYAWSSSEDFGT 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	64	38.8	161	4	US-09-270-767-42771	Sequence 42771, A
2	57.5	34.8	395	4	US-09-270-767-43336	Sequence 43336, A
3	55	33.3	352	2	US-08-472-534-6	Sequence 6, Appli
4	55	33.3	378	4	US-09-583-110-4647	Sequence 4647, Ap
5	55	33.3	379	4	US-09-107-433-3487	Sequence 3487, Ap
6	55	33.3	636	3	US-09-564-805-237	Sequence 237, App
7	54.5	33.0	522	3	US-09-142-732-2	Sequence 2, Appli
8	54.5	33.0	522	4	US-08-945-826-2	Sequence 2, Appli
9	54.5	33.0	522	4	US-09-197-503-2	Sequence 2, Appli
10	54	32.7	404	4	US-09-949-016-11198	Sequence 11198, A
11	54	32.7	419	4	US-09-252-991A-23245	Sequence 23245, A

12	52	31.5	458	5	PCT-US96-00994-4	Sequence 4, Appli
13	50	30.3	432	3	US-09-306-595C-8	Sequence 8, Appli
14	50	30.3	432	4	US-09-925-388-8	Sequence 8, Appli
15	50	30.3	446	4	US-09-949-016-10702	Sequence 10702, A
16	49.5	30.0	273	4	US-09-328-352-6316	Sequence 6316, Ap
17	49.5	30.0	371	2	US-08-442-809A-76	Sequence 76, Appl
18	49	29.7	23	1	US-08-004-139B-35	Sequence 35, Appl
19	49	29.7	23	2	US-08-811-492-35	Sequence 35, Appl
20	49	29.7	23	5	PCT-US96-10545A-35	Sequence 35, Appl
21	49	29.7	114	4	US-09-634-238-280	Sequence 280, App
22	48.5	29.4	141	2	US-08-345-321-10	Sequence 10, Appl
23	48.5	29.4	334	4	US-09-248-796A-16366	Sequence 16366, A
24	48.5	29.4	504	4	US-09-162-017-2	Sequence 2, Appli
25	48.5	29.4	521	4	US-08-945-826-6	Sequence 6, Appli
26	48.5	29.4	521	4	US-09-197-503-6	Sequence 6, Appli
27	48.5	29.4	997	3	US-09-369-364A-7	Sequence 7, Appli
28	48.5	29.4	1970	4	US-09-538-092-1005	Sequence 1005, Ap
29	48	29.1	128	4	US-09-270-767-34484	Sequence 34484, A
30	48	29.1	128	4	US-09-270-767-49701	Sequence 49701, A
31	48	29.1	139	4	US-09-050-739-68	Sequence 68, Appl
32	48	29.1	201	4	US-09-270-767-35706	Sequence 35706, A
33	48	29.1	201	4	US-09-270-767-50923	Sequence 50923, A
34	48	29.1	241	4	US-09-270-767-40578	Sequence 40578, A
35	48	29.1	241	4	US-09-270-767-55794	Sequence 55794, A
36	48	29.1	306	2	US-08-824-707-2	Sequence 2, Appli
37	48	29.1	320	4	US-09-248-796A-17463	Sequence 17463, A
38	48	29.1	979	4	US-09-538-092-990	Sequence 990, App
39	47.5	28.8	177	4	US-09-328-352-6964	Sequence 6964, Ap
40	47.5	28.8	521	4	US-08-945-826-4	Sequence 4, Appli
41	47.5	28.8	521	4	US-09-197-503-4	Sequence 4, Appli
42	47	28.5	239	4	US-09-134-000C-5005	Sequence 5005, Ap
43	47	28.5	239	4	US-09-248-796A-27281	Sequence 27281, A
44	47	28.5	247	4	US-09-270-767-46548	Sequence 46548, A
45	47	28.5	263	3	US-09-159-106-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

```

US-09-270-767-42771
; Sequence 42771, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42771
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid

```

US-09-270-767-42771

Query Match 38.8%; Score 64; DB 4; Length 161;
Best Local Similarity 47.6%; Pred. No. 0.49;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GSSGFGPYVAHGGYSGYEYAW 21
|: |:| | | |:| |:| :
Db 105 GAGGYGDYYAGGYNGYDYG 125

RESULT 3

US-08-472-534-6

; Sequence 6, Application US/08472534
; Patent No. 5919620
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard R
; APPLICANT: Martin, Denis
; TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,534
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr, James F
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: Biovac-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; TELEX: 14-8367
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-472-534-6

Query Match 33.3%; Score 55; DB 2; Length 352;

Best Local Similarity 43.5%; Pred. No. 17;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GSSGFGPYVAHGGYSGYEYAWSS 23
|: ||| : ||: |:| :||
Db 79 GAGGFGGFNGAGGFGGFEDIFSS 101

Search completed: November 8, 2005, 22:05:00
Job time : 43 secs